

Radiologie als onderdeel van het stationsexamen klinische vaardigheden

I.A.H. van den Berk, J.M.M. van de Ridder, J.P.J. van Schaik

Samenvatting

Inleiding: In een stationsexamen worden klinische vaardigheden gestandaardiseerd en contextgerelateerd getoetst. In vergelijking met een schriftelijke toets is vanwege de inzet van verschillende casus en beoordelaars de test minder betrouwbaar. In de preklinische fase van het medisch curriculum van het Universitair Medisch Centrum Utrecht wordt radiologie getoetst in een vaardigheidstoets. De vaardigheidstoets is een als een medisch consult opgezet stationsexamen. Om een indruk te krijgen van de betrouwbaarheid van het station Radiologie is het station geanalyseerd.

Methode: Er zijn vier vragen geformuleerd:

- Hoe is de interne consistentie van de items binnen het station Radiologie?
- Hoe is de score op het station Radiologie in vergelijking met de gehele toets?
- Wat zijn de scoreverschillen tussen de casus?
- Wat zijn de scoreverschillen tussen de beoordelaars?

De in 2004 bij tweedejaars studenten geneeskunde afgenomen vaardigheidstoets werd geanalyseerd.

Resultaten: Van de in totaal 265 deelnemers aan de vaardigheidstoets hebben 93 studenten de stationstoets radiologie afgelegd. De Cronbach alpha voor het station Radiologie is 0,92. De gemiddelde score op het station Radiologie is 3,8 (0,87). De gemiddelde score op de toets zonder radiologie is 3,9 (0,32). De range van de gemiddelde score tussen de zes casus is 0,5 (3,6-4,1). De range van de gemiddelde score tussen de vijf beoordelaars is 1,0 (3,3-4,3).

Discussie en conclusie: De interne consistentie van de items binnen het station Radiologie is goed. De gemiddelde score op het station Radiologie is vergelijkbaar met die op de andere stations. De range van de score tussen de verschillende casus is vrij gering. De range van de score tussen de beoordelaars is duidelijk groter. In de toekomst wordt geprobeerd dit verschil te verkleinen door een betere voorbereiding van de beoordelaars en door het gebruik van een globale scorerichtlijn tijdens de test. De grootste meerwaarde van het stationsexamen voor het toetsen van radiologische klinische vaardigheden zit in het interactief toetsen van de studenten. Op deze manier kan het interpreteren van radiologische onderzoeken optimaal getest worden. (Berk IAH van den, Ridder JMM van de, Schaik JPJ van. Radiologie als onderdeel van het stationsexamen klinische vaardigheden. Tijdschrift voor Medisch Onderwijs 2008;27(1):22-29.)

Inleiding

Het stationsexamen is ontwikkeld voor het toetsen van klinische vaardigheden.¹ Tijdens het examen wordt geprobeerd studenten zo objectief mogelijk te testen in een situatie die overeenkomt met de

praktijk. Studenten voeren in een station onder gestandaardiseerde condities een opdracht uit en worden daarbij geobserveerd. Hierbij wordt een praktijksituatie nagebootst. Met behulp van gestructureerde beoordelingscriteria en gestan-

daardiseerde casuïstiek wordt de student zo objectief mogelijk beoordeeld. De betrouwbaarheid van deze toetsvorm is gevoeliger dan een schriftelijke test; daarom zijn meerdere stations nodig.^{2,4} De inhoudspecificiteit van de casus speelt hierbij een belangrijke rol. Dit vergroot de kans dat een student hetzelfde cijfer haalt bij een andere casus.^{2,3}

In de preklinische fase (het eerste en tweede jaar) van het medisch curriculum van het Universitair Medisch Centrum Utrecht (UMCU) maakt radiologie deel uit van het praktisch lijnonderwijs. Bij het interpreteren van radiologische onderzoeken worden radiologische bevindingen gecorreleerd met de klinische context. Dit maakt radiologie tot een klinisch vak dat past binnen het op de praktijk gerichte praktisch lijnonderwijs. In twaalf practica van twee uur wordt de studenten aan de hand van casuïstiek geleerd hoe ze de meest voorkomende radiologische onderzoeken systematisch moeten beoordelen. Met behulp van de radiologische onderzoeken leert de student de toegepaste anatomie en krijgt hij/zij inzicht in de toegepaste onderzoeksmethode en techniek. Het praktisch lijnonderwijs wordt aan het einde van het studiejaar getoetst in een stationsexamen: de vaardigheidstoets.

Daar er veel variabelen meespelen in het stationsexamen (de inzet van zowel verschillende casus als beoordelaars), ontstond de vraag wat de waarde is van deze toetsvorm. Uit literatuuronderzoek blijkt dat er over het toetsen van radiologische klinische vaardigheden in een stationsexamen nog niet veel bekend is.⁵⁻⁶ Met behulp van de analyse van de toetsresultaten van het station Radiologie in de vaardigheidstoets is geprobeerd hier inzicht in te krijgen.

Methode

Studentenpopulatie

In 2004 is de vaardigheidstoets van twee-dejaars studenten geneeskunde geanalyseerd.

Materiaal

De vaardigheidstoets van het praktisch lijnonderwijs in het UMCU is opgezet als een medisch consult met vier stations: anamnese, lichamelijk onderzoek, aanvullend onderzoek en verslaglegging. Radiologie wordt mondeling getoetst in het station 'aanvullend onderzoek'. Een deel van het studentencohort krijgt het station Radiologie; de overige studenten worden getoetst op een klinische vaardigheid van een ander vakgebied (bijvoorbeeld het hechten van een wond bij Chirurgie). De studenten weten vooraf niet op welk vakgebied ze in het station 'aanvullend onderzoek' worden getoetst.

Het station Radiologie bestaat uit drie onderdelen: *interpretatie*, *anatomie* en *onderzoeksmethode en techniek*. Naar aanleiding van een casusbeschrijving vraagt de student onderzoeken aan en beoordeelt deze onderzoeken vervolgens (interpretatie). Aan de hand van de gekozen onderzoeken wordt de anatomische kennis getoetst en worden er vragen gesteld over de gebruikte onderzoeksmethode en techniek. De beoordeling van de student vindt plaats met behulp van een beoordelingslijst (figuur 1). De lijst bestaat uit vier hoofditems: interpretatie, anatomie, onderzoeksmethode en techniek en algemene indruk. De eerste drie hoofditems zijn onderverdeeld in respectievelijk vier, twee en twee subitems.

De vier subitems van het hoofditem *interpretatie* zijn: aanvragen eenvoudig radiologisch onderzoek aan de hand van casusbeschrijving, beoordeling eenvoudig radiologisch onderzoek, aanvragen complex radiologisch onderzoek aan de hand

Toets PLO radiologie		Naam student: _____				
Casus 2A		Colifort nr: _____				
Man 55 jaar. Bij toeval medaan in de buik pulserende drukprijke zwelling gevonden						
1. Welk radiologisch onderzoek vraag je aan? Onderzoek: Indicatie:	1	2	3	4	5	
2. Geef Echo: Bespreek bevindingen en diagnose. (*) Lumen Thrombus Verrekening Proces II Diagnose	1	2	3	4	5	
3. Welk aanvullend radiologisch onderzoek vraag je aan? Onderzoek: Indicatie:	1	2	3	4	5	
4. Geef CT-aorta: Bespreek bevindingen en diagnose. (*) Lumen Thrombus Verrekening Relatie nierenarteriën en aa. iliaac Diagnose	1	2	3	4	5	
Beoordeling interpretatie:	1	2	3	4	5	
5. Laaf samen: Rechter nier V. cava inferior Lever Galblaas M. proes	1	2	3	4	5	
6. Wje aar en jvat benoemen: Pancreas A. mesenterica superior Mesenterium Linker vena renalis M. iliacus	1	2	3	4	5	
Beoordeling anatomie:	1	2	3	4	5	
7. Geef brief inleiding echografie: Leg uit: Principe echografie Ultrageluidsgolven	1	2	3	4	5	
8. Bespreek reflectie ultrageluidsgolven: Acoustische impedantie	1	2	3	4	5	
Beoordeling onderzoeksmethode en techniek:	1	2	3	4	5	
Algemene indruk	1	2	3	4	5	
1 = zeer onvoldoende 2 = onvoldoende 3 = matig 4 = voldoende 5 = goed		Init. beoordelaar: _____ Citeer Casus 2 en 4, blad 2 van 3 1/1, naar 4 1/1 Citeer: Glas: Inleiding Echografie				
Vragen met een (*) hebben een drie dubbele weging voor dat onderdeel.						

Figuur 1. Beoordelingslijst van casus A) Aneurysma van de abdominale aorta.

van de radiologische bevindingen, beoordelen complex radiologisch onderzoek. De twee subitems bij het hoofditem *anatomie* zijn: aanwijzen en benoemen van anatomische structuren. De twee subitems van het hoofditem *onderzoeksmethode en techniek* zijn twee vragen over één van de twee tijdens het station gebruikte radiologische onderzoeksmethoden.

De hoofditems worden gescoord aan de hand van de subitems. De subitems van het hoofditem *interpretatie* hebben een verschillende weging. Het goed kunnen beoordelen van het onderzoek weegt zwaarder dan de keuze van het onderzoek aan de hand van de klinische context. Als laatste wordt een score gegeven voor de algemene indruk die de student op de beoordelaar maakt.

De studenten worden beoordeeld op een 5-puntsschaal aan de hand van een globale scoringslijst (1: zeer onvoldoende; 2: onvoldoende; 3: matig; 4: voldoende; 5: goed). Tijdens de toets zijn vijf beoordelaars ingezet. Alle beoordelaars zijn arts-assistenten radiologie die betrokken zijn geweest bij het geven van de practica radiologie in het praktisch lijnonderwijs. De arts-assistenten hebben dus inzicht in het onderwijs dat voorafging aan het stationsexamen en daardoor affiniteit met het afnemen van de toets. Zij hebben tijdens dit practicum een groep van 12 studenten les gegeven. Met een studentencohort van 256 studenten is de kans klein dat zij tijdens het afnemen van de toets hun 'eigen' studenten treffen.

Er zijn in het station Radiologie zes gestandaardiseerde casus gebruikt. Alle zes casus zijn afkomstig uit de practica radiologie van het praktisch lijnonderwijs. De onderwerpen van de casus zijn A) aneurysma van de abdominale aorta, B) bronchuscarcinoom, C) emfyseem, D) pancreaskopcarcinoom, E) ruptuur van de mediale meniscus van de knie en F) oeso-

faguscarcinoom. De onderwerpen lopen qua inhoud uiteen maar de opzet van de casus is elke keer hetzelfde. Het gebruik van verschillende casus voorkomt dat studenten informatie aan elkaar kunnen doorgeven. De casus wisselen na het toetsen van zes studenten. Dit geldt ook voor de inhoud van het station 'Anamnese' en 'Lichamelijk onderzoek'. De beoordelaars hebben geen vaste casus.

Procedure en analyse

Om een indruk te krijgen van de betrouwbaarheid van het stationsexamen Radiologie zijn vier vragen geformuleerd:

- Hoe is de interne consistentie van de items binnen het station Radiologie?
- Hoe is de score op het station Radiologie in vergelijking met de gehele toets?
- Wat zijn de scoreverschillen tussen de casus?
- Wat zijn de scoreverschillen tussen de beoordelaars?

De toetsresultaten zijn met behulp van het computerprogramma SPSS geanalyseerd. Voor het beoordelen van de interne consistentie binnen het station is de Cronbach alpha bepaald.⁷ De gemiddelde score op het station Radiologie is vergeleken met de gemiddelde score op de toets zonder radiologie, die bepaald werd uit de toetsresultaten van de 93 studenten die hadden deelgenomen aan het station Radiologie. De verschillen in score tussen de casus zijn bepaald door de gemiddelde score van de studenten op de verschillende casus met elkaar te vergelijken. De scoreverschillen tussen de beoordelaars zijn bepaald door de door de beoordelaars gegeven gemiddelde scores te vergelijken. Op deze manier is geprobeerd een globale indruk te krijgen van de invloed van het inzetten van verschillende casus en beoordelaars tijdens de toets. De aantallen zijn klein en de verdeling van casus over de be-

Tabel 1. De verdeling van de casus over de beoordelaars*.

Beoordelaars	Casus A 3,9	Casus B 3,8	Casus C 3,6	Casus D 3,9	Casus E 3,6	Casus F 4,1
B1 4,1	8	0	8	8	0	0
B2 4,3	0	4	0	0	6	11
B3 3,3	5	0	6	6	1	8
B4 3,6	4	0	0	0	3	0
B5 3,8	0	10	3	2	0	0

* Onder de casus staat de gemiddelde score op de casus. Naast de beoordelaar staat de gemiddelde door de beoordelaar gegeven score.

oordelaars is willekeurig en niet gelijkmatig (tabel 1). Daar het aantal keren dat een beoordelaar of casus is ingezet niet gelijk is (tabel 2 en 3), is het onmogelijk om een statistische analyse toe te passen. Ook kan de generaliseerbaarheidstoets niet toegepast worden omdat elke student één casus krijgt en gezien wordt door één beoordelaar.⁸

Resultaten

Van de 265 studenten die in 2004 hebben deelgenomen aan de vaardigheids-toets hebben 93 studenten bij het onderdeel 'aanvullend onderzoek' het station Radiologie gehad. De op de vier hoofd-items gebaseerde Cronbach alpha is 0,92. De gemiddelde score op het station Radiologie is 3,8 (sd 0,87). De gemiddelde score op de toets zonder radiologie is 3,9 (sd 0,32). De range van de gemiddelde

score tussen de zes casus is 0,5 (laagste score 3,6; hoogste score 4,1 – zie tabel 2). De range van de gemiddelde score tussen de vijf beoordelaars is 1,0 (laagste score 3,3; hoogste score 4,3 – zie tabel 3).

De door de beoordelaars gegeven gemiddelde scores zijn vergeleken met de gemiddelde score van de studenten op de toets

Tabel 2. De gemiddelde score van de zes casus.

Casus	n	Gemiddelde vier hoofditems (sd)
A	17	3,9 (0,58)
B	14	3,8 (0,76)
C	16	3,6 (0,94)
D	16	3,9 (0,91)
E	9	3,6 (0,99)
F	21	4,1 (1,0)

Tabel 3. De gemiddelde score van de vijf beoordelaars.

Beoordelaars	n	Gemiddelde score radiologie (sd)	Gemiddelde score toets zonder radiologie (sd)
B1	24	4,1 (0,67)	4,0 (0,26)
B2	21	4,3 (0,93)	3,9 (0,31)
B3	26	3,3 (0,82)	3,7 (0,33)
B4	7	3,6 (0,79)	3,7 (0,34)
B5	15	3,8 (0,70)	4,0 (0,29)

zonder radiologie. De Levene's statistics tussen deze twee groepen is niet significant. De hypothese dat de varianties tussen de groepen verschillen mag niet worden verworpen. De ANOVA geeft een significante F-waarde (0,009). De hypothese dat de toetsscores gelijk zijn tussen de beoordelaars moet worden verworpen.

Discussie

Het doel van de analyse van de toetsresultaten van het station Radiologie is het krijgen van inzicht in de betrouwbaarheid van het station. Wat is de invloed van de inzet van verschillende casus en van verschillende beoordelaars tijdens de toets?

Onze analyse van het station Radiologie richt zich vooral op twee basale eisen van een goed station: de interne consistentie van de schaal (alpha) en de vergelijkbaarheid van de scores met die van andere toetsstations. De interne consistentie van het station is een maat voor de betrouwbaarheid van het station. De Cronbach alpha is goed indien $> 0,7$ is. Een Cronbach alpha van 0,92 komt overeen met een hoge interne consistentie tussen de vier hoofditems. De gemiddelde score op het station Radiologie (3,8) is vergelijkbaar met de gemiddelde score op de toets zonder radiologie (3,9). De standaarddeviatie van scores op het station Radiologie (0,87) is groter dan die van de scores op de toets zonder radiologie (0,32). Voor een groot deel kan dit verklaard worden door het grotere aantal items waaruit de score op de toets zonder radiologie opgebouwd is. Het verschil in gemiddelde score tussen de casus is 0,5. Het verschil in gemiddelde score tussen de beoordelaars is 1,0 en opvallend groter in vergelijking met die tussen de casus.

Als naar de verdeling van de casus over de beoordelaars wordt gekeken (tabel 1) is er geen directe relatie tussen de door de beoordelaar gegeven gemiddelde score en

de gemiddelde score op de beoordeelde casus. Als de gemiddelde score door de beoordelaars wordt vergeleken met de gemiddelde score van hun studenten op de toets zonder radiologie laat een ANOVA een significant verschil zien. Het is dus mogelijk dat de éne beoordelaar betere studenten heeft gehad dan de andere beoordelaar (tabel 3) en daarmee is het onmogelijk om een harde uitspraak te doen over de gevonden verschillen in range tussen casus en beoordelaars.

In de toetsen die na 2004 zijn afgenomen is geprobeerd de eenduidigheid tussen de beoordelaars te vergroten. Voorafgaande aan de toets worden de casus en manier van scoren uitgebreid met de beoordelaars besproken. Met behulp van een globale scorerichtlijn is geprobeerd het scoren eenduidiger te maken. Uit de literatuur blijkt dat een globale scorerichtlijn een betere beoordeling geeft van het niveau van een student dan een uitgebreide checklist.^{2-3 9-11}. Er is gekozen voor een grove scoremaat in procenten die gebruikt kan worden voor alle sub- en hoofditems. Score 1 $< 10\%$: goed; score 2 $< 55\%$: goed; score 3 tussen de 55-70%: goed; score 4 70%: goed of alles goed met hulp; score 5 100%: direct goed. De ervaringen met deze scorerichtlijn zijn positief. Het geeft de beoordelaars ondersteuning in het geven van een onderbouwde en consistente beoordeling.

Onze ervaringen met het geven van practica en het toetsen van radiologie in het kader van het praktisch lijnonderwijs zijn zeer positief. Het beoordelen van een radiologisch onderzoek is een klinische vaardigheid. Om dit goed te kunnen doen is inzicht nodig in de klinische situatie van de patiënt, kennis van de anatomie en kennis van de gebruikte onderzoeksmethode en techniek. Het is tevens belangrijk om te weten wat in een bepaalde klinische context het onderzoek van eerste keuze is.

Het stationsexamen biedt de mogelijkheid om de studenten interactief te toetsen. In vergelijking met een schriftelijke toets wordt meer inzicht verkregen in de wijze waarop een student tot een bepaald antwoord komt. Het is mogelijk om tijdens de toets de student een hint te geven in welke richting het goede antwoord gezocht moet worden. Over het geven van hints zijn voorafgaande aan de toets met de beoordelaars duidelijke afspraken gemaakt. Het geven van een hint gaat direct ten koste van de score van de student.

Door het contact met de studenten geeft de toets een directe feedback op het gegeven onderwijs.

Een nadeel is dat de toetsvorm tijdsintensief is en het daardoor niet mogelijk is alle studenten te toetsen. De studenten die het station Radiologie krijgen worden slechts aan de hand van één casus en door één beoordelaar getoetst. De inhoudsspecificiteit van de casus en de observatie door één beoordelaar beperken de betrouwbaarheid van de toets.²⁻³ Het valt te overwegen om aanvullend aan het stationsexamen alle studenten in een schriftelijke toets op radiologische kennis te toetsen. In combinatie met het stationsexamen kan dit de betrouwbaarheid verhogen.⁴

Conclusie

De interne consistentie van de vier hoofd-items binnen het station Radiologie is goed. De gemiddelde uitslag op het station Radiologie is vergelijkbaar met de gemiddelde uitslag van de toets zonder radiologie. De spreiding tussen de beoordelaars is groter in vergelijking met de casus. Naar aanleiding van deze bevindingen is met behulp van een intensievere voorbereiding van de beoordelaars en een globale scoringslijst geprobeerd de spreiding te verlagen. De meerwaarde van het stationsexamen voor het toetsen van radiologische klinische vaardigheden zit in het interactief toetsen van de studenten tijdens de toets. Op deze manier kan het interpreteren van radiologische onderzoeken, het correleren van de radiologische bevindingen met de klinische context, optimaal getest worden.

ding te verlagen. De meerwaarde van het stationsexamen voor het toetsen van radiologische klinische vaardigheden zit in het interactief toetsen van de studenten tijdens de toets. Op deze manier kan het interpreteren van radiologische onderzoeken, het correleren van de radiologische bevindingen met de klinische context, optimaal getest worden.

Literatuur

1. Harden RM, Stevenson M, Wilson Downie W, Wilson GM. Assessment of clinical competence using objective structured examination (OSCE). *Br Med J* 1975;1:445-51.
2. Metz JCM, Scherpier AJJA, Vleuten CPM van der. Medisch onderwijs in de praktijk. 7.6 Toetsing van medische competenties. Het stationsexamen. Van Gorcum & Comp 1995;202-7.
3. Vleuten CPM van der, Driessen EW. Hoger onderwijs Praktijk Toetsing in probleemgestuurd onderwijs. 4.4.1 Toetsing van beroepsspecifieke vaardigheden. Het stationsexamen. Wolters-Noordhoff 2000;47-53.
4. Verhoeven BH, Hamers JGHC, Scherpier AJJA, Hoogenboom RJ, Vleuten CPM van der. The effect on reliability of adding a separate written assessment component to an objective structured clinical examination. *Med Educ* 2000;34:525-9.
5. Morag E, Lieberman G, Volkan K, Shaffer K, Novelline R, Lang E. Clinical Competence Assessment in Radiology. *Acad Radiol* 2001;143:795.
6. Novelline RA, Squire LF, Whitley JE. Written and practical testing of medical students in radiology. *Radiology* 1982;143:795-7.
7. <http://www.cito.nl/share/beglist1.htm#ALPHA>
8. Vleuten CPM van der, Meier AHP. De schatting van de reproduceerbaarheid van meetgegevens met behulp van generaliseerbaarheidstheorie: ook in het fysische domein? *Klinische Fysica* 1:4-10.
9. Regehr G, MacRae H, Reznick RK, Szalay D. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. *Ac Med* 1998;73:993-7.
10. Newble D. Techniques for measuring clinical competence: objective structured clinical examination. *Med Educ* 2004;38:199-203.
11. Wilkinson TJ, Frampton CM, Thompson-Fawcett M, Egan T. Objectivity in objective structured clinical examinations: Checklist are no substitute for examiner commitment. *Ac Med* 2003;2:219-23.

De auteurs:

Mw. drs. I.A.H. van den Berk was tijdens het onderzoek aios Radiologie.

Mw. drs. J.M.M. van de Ridder is aio Onderwijsinstituut.

Dr. J.P.J. van Schaik is radioloog. Tijdens het onderzoek waren allen werkzaam in het UMC Utrecht.*

* vanaf 1 november 2007 is mw. van den Berk werkzaam als radioloog in het LUMC.

Correspondentieadres:

Inge van den Berk, LUMC, afdeling Radiologie, Postbus 9600, 2300 RC Leiden.

Email: I.A.H.van_den_Berk@LUMC.nl

Belangenconflict: geen gemeld.

Financiële ondersteuning: geen gemeld.

Summary

Introduction: Objective structured clinical examinations (OSCE) assess clinical competence in a standardised and context related manner. Compared with written tests, OSCES are more susceptible to reliability errors because of the use of multiple cases and multiple examiners. In the preclinical phase of the medical curriculum of University Medical Centre Utrecht, an OSCE is organised as a medical consultation. We evaluated the radiology station.

Methods: We sought to answer four questions:

- What is the internal consistency of the items of the radiology station?
- How do the scores on the radiology station compare with the total scores on the other stations?
- How do different cases differ in scores?
- What are the differences in scores, if any, between the examiners?

We analysed the OSCE results of second year medical students in 2004.

Results: In 2004, 265 students participated in the OSCE. Of these, 93 students were tested in the radiology station. The alpha coefficient of this station was 0.92; the mean score was 3.8 (0.87). The total mean score, excluding the radiology station, was 3.9 (0.32) and the mean scores on the six cases ranged from 3.6 to 4.1. The mean scores of the five examiners ranged between 3.3 and 4.3.

Conclusion: The internal consistency of the items of the radiology station is good and the mean score is comparable with that of the other stations, with a fairly limited range of scores between cases and a wider range between the examiners. We aim to reduce this difference by preparing examiners better for their task and by introducing a global rating list. An advantage of the OSCE as an instrument to test clinical competence in radiology is its interactive nature. An OSCE enables optimal assessment of interpretation of radiological investigations. (Berk IAH van den, Ridder JMM van de, Schaik JPI van. A radiology station in an objective structured clinical examination. An evaluation of the radiology station in the skills test of the undergraduate medical curriculum of Utrecht University Medical Center. Dutch Journal of Medical Education 2008;27(1):22-29.)